JANUARY/FEBRUARY 1984
Volume 12 Number 1

I.P. Sharp newsletter

in this issue

Institute of International Finance 1
New Local Access Numbers 2
SHARP APL/PC Transfer Facility..... 3
I. P. Sharp Data Bases: as Easy as 1–2–3. 3
Enhancements to EASY, the Econometric Analysis System..... 4
Petroflash! Oil Prices Instantly at Your Fingertips..... 6
DeWitt Petrochemical Newsletters in INFOMAGIC..... 7
1982CFD Data Base Targets Markets for Financial Services 8
I. P. Sharp's First Distributed Software Symposium 10
1984 Users Meeting 10
Bookends.....11

Technical Supplement 48
Making Things Fit, Part 3 .....T1

Institute of International Finance

I. P. Sharp Associates is to supply computer services around the world to the 180 member banks of the Institute of International Finance (IIF), headquartered in Washington, D.C.

The IIF is committed to providing the world's financial community with improved availability and quality of information about the economic situation and evolving debt position of borrower countries. This information is for the member banks' use in individually assessing credit to public and private sectors in the international markets.

The IIF member banks collectively hold over three-quarters of the world's total bank debt to non-OECD countries. Membership in the IIF is open to any bank with cross-border exposure.

Currently the IIF is focusing on the collection, analysis, and dissemination of data. In

customer application

. . . Institute of International Finance

particular, the IIF intends to fill in the gaps in existing data collection efforts by providing up-to-date information in a more timely fashion. The IIF is placing a special emphasis on improving the information currently available to private commercial banks from published sources.

In order to accomplish this aim, the Country Evaluation System was jointly developed by the IIF and I. P. Sharp. This system stores IIF economic data in a private data base on the I. P. Sharp timesharing service, where it is instantly available to any member bank. In most cases, with a telephone call placed locally to connect them to the I. P. Sharp Network, a member bank can selectively display any of the data available from the IIF, and make credit judgements accordingly.

When the data base is released to the IIF member banks early this year, they will have instant access to data for the past ten years, plus estimates for the current year, and in some cases, projections, on about 75 different items for some 20 developing nations. This data covers central government budget, balance of payments, long-term and total debt, and key economic indicators.

Country Evaluation System

The Country Evaluation System is easy to learn and to use. It was designed so that member banks can retrieve salient facts from the Institute's data base with the minimum of effort. Reports are pre-formatted; simple, English commands produce reports and graphs on most terminals.

The Institute of International Finance is not geared just to the world's largest international banks, who often have their own economic and financial analysts on staff plus a means of gathering data through their branches in the countries where they do business. The IIF represents a big advance to the regional banks, who hold a significant proportion of the debts owed by lesser developed countries. As such, they represent an important factor in international lending.

By providing the regional banks with more information, gathered and analysed by an independent body, the IIF will enable regional banks to make better and surer decisions.

Through the IIF's use of the I. P. Sharp Network, all IIF member banks will also have online access to over 100 public data bases. Those of specific interest to IIF members include the World Bank Debt Tables, the Bank for International Settlements, the IMF's International Financial Statistics, and the United Nations Commodity Trade Statistics. These data bases provide supplementary information to the main data of the Country Evaluation System.

As members of the Institute of International Finance, the world's international lenders, whatever their size, now have more up-to-date information for analysis of country creditworthiness right at their fingertips. ■

communication news

New Local Access Numbers

Minneapolis

120 cps 612 338–4918

Philadelphia

30 cps 215 238–1005
120 cps 215 238–1004

Phoenix

30 cps 602 274–2188
120 cps 602 274–2189

Regina

120 cps 306 359–0421

SHARP APL/PC Transfer Facility

I. P. Sharp's Application Software Group in Toronto is developing a generalized upload and download facility, scheduled for release early in 1984, to link SHARP APL on a mainframe to SHARP APL running on an IBM PC. With this facility you can transfer APL objects, or the entire contents of your workspaces and files, between the mainframe and an IBM PC, or an IBM PC local area network.

This first release uses asynchronous communications between the IBM PC and a mainframe. The communications interface to APL on the IBM PC is performed by sharing variables with an auxiliary processor. Transmission error detection and correction is built into the transfer functions.

The transfer facility performs all necessary translations to convert APL objects into transmittable characters and back again. It distinguishes among boolean, integer, real, and complex numbers, and enclosed arrays, functions, and packages. All transferred APL objects conform to their original data type.

You have the choice of using the transfer facility in one of two modes—as a *closed* system where you interactively specify files and workspaces to be transferred, or as an *open* system where you use a set of APL utilities to send and receive data between the IBM PC and a mainframe.

Closed system

The closed system uses a full-screen interface. With the closed version, you use the IBM PC as an intelligent terminal to sign on to a mainframe. You then enter the direction of the transfer (uploading from the IBM PC, or downloading to the IBM PC), and the names of origin and target files, and workspaces. You use the closed system to distribute applications to remote sites. Using the closed system, you can take advantage of the superior response time on the mainframe to develop, test, and debug software to be run on the IBM PC.

Open system

The open system is a set of APL utilities for programmers to build APL applications which share data between a mainframe and the IBM PC. This allows you to exchange individual APL objects between the IBM PC and a mainframe under program control. Using the transfer utilities, you can use the IBM PC as the front end of an application running on the mainframe. Alternatively, you can build applications running on several remote IBM PCs which transmit data to a central mainframe for data collection and consolidation. ■

Peter Airs, Toronto

I. P. Sharp Data Bases: as Easy as 1–2–3

The personal computer has greatly extended your computing power. As the user of a personal computer, you can design your own information system instead of specifying it for someone else to design for you. But what good is an information system without the essential ingredient—information?

Certainly you could enter the data from a printed source, with the inevitable problems of accuracy and reliability, not to mention the time wasted in this duplication of effort. Alternatively, you can merge the old technology with the new: by using the I. P. Sharp public data bases through the I. P. Sharp timesharing service and the Lotus Development Corporation's Lotus 1–2–3 spreadsheet system on your personal computer.

You use the public data bases as the *data source* and 1–2–3 as the *data manager*. Ideally, the link between the two should be as transparent as possible: you start your 1–2–3 session, and communicate with the I. P. Sharp timesharing service to move data down to your spreadsheet.

Unfortunately, the current version of 1–2–3 cannot do this. Until it can, you have to provide your own communications link between the I. P. Sharp data bases and your spreadsheet. This involves the following steps:

. . . I. P. Sharp Data Bases

1. Sign on to the I. P. Sharp timesharing service, using a communications package on your personal computer, such as I. P. Sharp's MICROCOMM.
2. Proceed with your data base retrieval session as though you were signed on to a normal "dumb" terminal.
3. At the point where you would normally display the accessed data, capture the display in a file on your personal computer.
4. Finally, sign off the I. P. Sharp timesharing service, and use 1–2–3 commands to bring up the captured data in a spreadsheet.

Sounds simple? It is. In previous *Newsletter* articles, you have read how you can use MICROCOMM to sign on to a SHARP APL system, using the IBM Personal Computer as an intelligent terminal on the I. P. Sharp Network. Then, using the *MICROPUT* command in MAGIC, you can store the transmitted data in Software Arts' data interchange format (DIF) files for later translation to a spreadsheet file. This is a simple, straightforward, standard way to transfer spreadsheet data.

1–2–3 supports DIF files, although they do have to be translated again, to a form readable by 1–2–3, before they can be used in a spreadsheet. Another way to convert data to the 1–2–3 spreadsheet format is the print file (PRN) format. This is a fast and effective means for transforming data into a 1–2–3 spreadsheet.

To transfer data from a data base to the PRN format, you must first load the appropriate MAGIC workspace (for example, 39 *MAGIC*), and set the desired timeframe. The data of interest can then be loaded to diskette as in the following example, which loads the high, low, close, and volume information for November 1983 for IBM as traded on the New York Stock Exchange.

```
      )LOAD 39 MAGIC
SAVED   4.31.83 12/20/83
      1 2 3 4 5 DAILY DATED AT 11 83
      AUTOLABEL
      )COPY 2056371 LOAD123
SAVED   16.15.49 12/12/83
      LOAD123 'NIBM' USSTOCKD 'HIGH,LOW,CLOSE,VOLUME'
```

Once the data elements have been transmitted to a PRN file, you can transfer them onto your spreadsheet by positioning the cell cursor to the top left corner of where you wish the data to go. Then use the 1–2–3 command **`/File Import Numbers`** to bring the data in. Simple as 1–2–3, isn't it? ■

Josh Levine, Dallas

application software

Enhancements to EASY, the Econometric Analysis System

EASY now includes many enhancements which, for the most part, have been suggested by its users. The new EASY features additional estimation procedures, improved and extended output capabilities, and easier input commands. Some of the changes are new, while others make it more convenient for you to use EASY.

A few of the more important enhancements are:

- Two-stage least squares
- Weighted least squares
- Application of various generalized least squares procedures to equations involving distributed lags
- An improved plot facility, including automatic application of SUPERPLOT
- Abbreviated input commands and a means of continuing lengthy input strings on more than one line
- An edit facility for correcting errors in, or making revisions to, input commands

You can find more information concerning all of these enhancements in EASY's *NEWS* system. The function *NEWS* in the workspace 39 *EASY* documents all significant changes to EASY since the publication of the *EASY User's Guide* (June 1982).

. . . Enhancements to EASY

To use the *NEWS* system, you simply load the workspace and type:

```
NEWS
```

A list of valid news item numbers prints immediately, followed by the prompt:

```
* WHICH ITEM NUMBER(S):
```

You then type the word *ALL*, or select one or more numbers from the list of valid item numbers.

A response of *S* or *D* to the subsequent prompt:

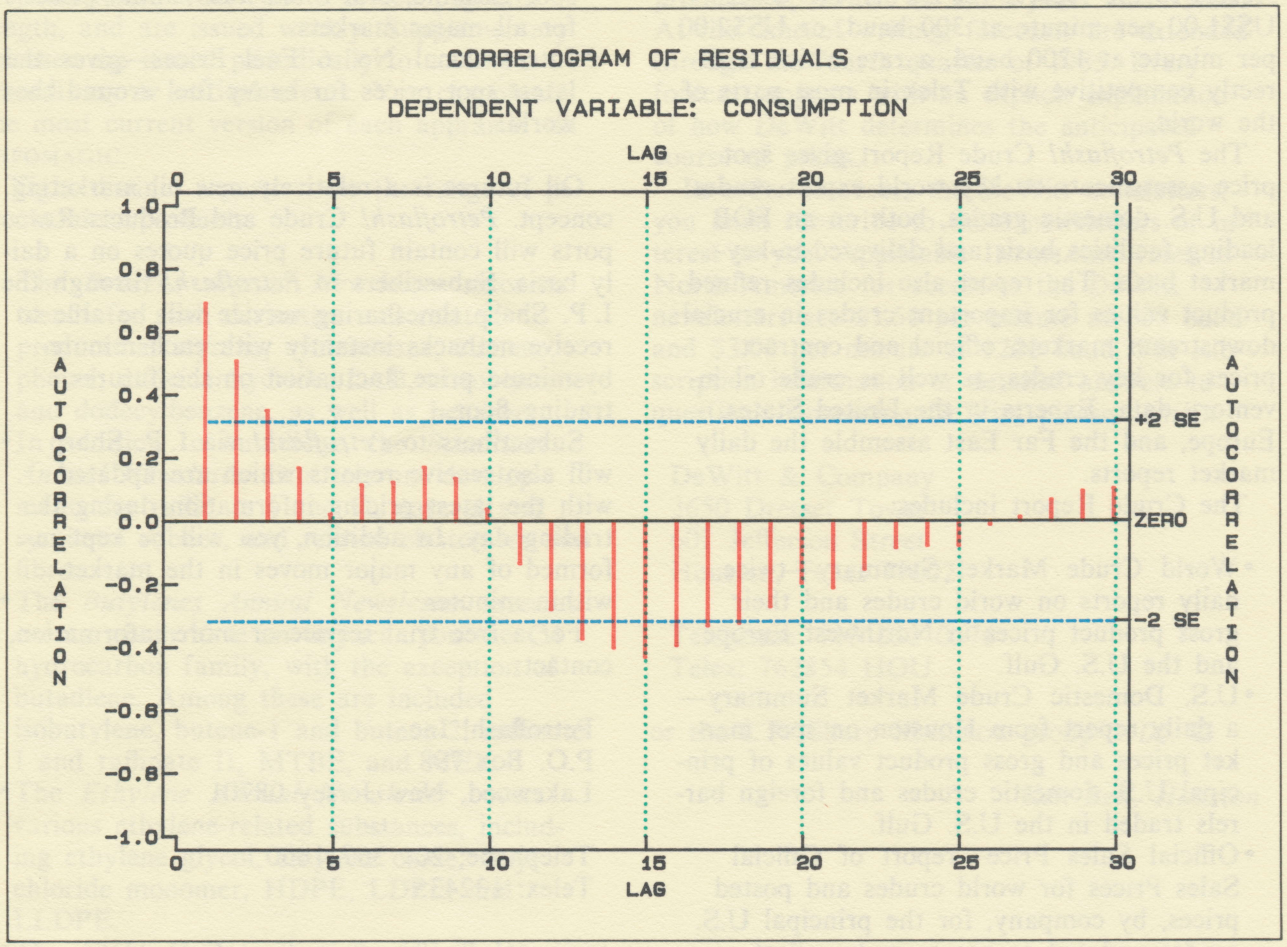
```
* SUMMARY OR DETAILED DESCRIBE? (S/D):
```

indicates whether you want headlines or full documentation, respectively, for each selected item.

The accompanying plot illustrates EASY's new plot capability. It was produced by executing the usual *PLOT RESIDUALS* command with a new EASY state setting in effect.

EASY will continue to evolve based on your requests. Please send your comments and suggestions to the MAILBOX group *EASYQ*. ■

Andrew North, Ottawa



Petroflash! Oil Prices Instantly at Your Fingertips

The publishers of *Petroleum Intelligence Weekly* (*PIW*) and *Oil Buyers' Guide* (*OBG*) now offer *Petroflash!*, a daily oil price information system, through INFOMAGIC on the I. P. Sharp timesharing service.

Combined, *PIW* and *OBG* have almost 35 years in oil reporting behind them. Both are accurate and respected trade publications in petroleum marketing.

Petroflash! comprises two daily reports—one on crude and the other on products. The subscription fee for either report is US$400 per year, or both reports for US$750 per year. Access to the reports via INFOMAGIC costs US$1.00 per minute at 300 baud, or US$2.00 per minute at 1200 baud, a rate which is directly competitive with Telex in most parts of the world.

The *Petroflash!* Crude Report gives spot price assessments on key world export crudes and U.S. domestic grades, both on an FOB loading facilities basis and delivered-to-key-market basis. The report also includes refined product values for important crudes in crucial downstream markets, official and contract prices for key crudes, as well as crude oil inventory data. Experts in the United States, Europe, and the Far East assemble the daily market reports.

The Crude Report includes:

- World Crude Market Summary—twice daily reports on world crudes and their gross product prices in Northwest Europe and the U.S. Gulf.
- U.S. Domestic Crude Market Summary—a daily report from Houston on spot market prices and gross product values of principal U.S. domestic crudes and foreign barrels traded in the U.S. Gulf.
- Official Sales Price—report of Official Sales Prices for world crudes and posted prices, by company, for the principal U.S. crudes updated as necessary to reflect changes.
- End-of-the-Week Review—evaluates the most widely traded crudes on the basis of their spot prices and product-yield-values.

The *Petroflash!* Products Report is a comprehensive appraisal of spot and contract product prices in the U.S., Caribbean, and Europe. The Products Report gives oil marketers a quick and accurate report of daily market activity in the major world trading centres. Anyone involved with buying or selling oil products will find this service invaluable.

The Products Report includes:

- U.S. Spot and Contract Product Prices—details the closely watched East Coast No. 6 fuel spot and contract prices.
- European Products Price Wrap-up—shows the market prices for both the European open and close of business.
- International Light Product Prices—lists spot gasoline, No. 2 oil, and naptha prices for all major markets.
- International No. 6 Fuel Prices—gives the latest spot prices for heavy fuel around the world.

Oil futures is a relatively new oil marketing concept. *Petroflash!* Crude and Products Reports will contain future price quotes on a daily basis. Subscribers to *Petroflash!* through the I. P. Sharp timesharing service will be able to receive netbacks instantly with each minute-by-minute price fluctuation on the futures trading floor.

Subscribers to *Petroflash!* via I. P. Sharp will also receive reports which are updated with the latest pricing information during the trading day. In addition, you will be kept informed of any major moves in the market within minutes.

For a free trial service or more information, contact:

Petroflash! Inc.
P.O. Box 798
Lakewood, New Jersey 08701

Telephone: 201 367–1600
Telex: 132435

or your I. P. Sharp representative, or the MAILBOX group *OILQ*. ■

DeWitt Petrochemical Newsletters in INFOMAGIC

DeWitt & Company, the widely-respected petrochemical consulting firm, produces a series of critical analyses and price forecasts for the petrochemical industry. Now 16 DeWitt petrochemical newsletters and price forecasts are available via INFOMAGIC on the I. P. Sharp timesharing service.

The DeWitt newsletters provide a comprehensive yet concise assessment of current developments (supply, demand, production, consumption, and context) and price trends for such chemical groupings as the paraffins, lower olefins, diolefins, aromatics and naphthenes. Reports range from two to fifteen pages in length, and are issued weekly. Supplementary releases are issued periodically to cover extraordinary or critical events. You can obtain the most current version of each appraisal via INFOMAGIC.

Eight distinct reports cover the range of petrochemical products:

- The *Benzene Annual Newsletter* recounts trends for such derivative and related products as styrene, cyclohexane, cumene, phenol, maleicanhydride, LAB, nitrobenzene and dodecylbenzene, as well as benzene.
- In addition to butadiene, the *Butadiene Annual Newsletter* lists relevant developments for raffinate I and II, natural and synthetic rubber, and related elastomers and fibres.
- The *Butylenes Annual Newsletter* assesses events which have an impact upon the C4 hydrocarbon family, with the exception of butadiene. Among these are included isobutylene, butene-1 and butene-2, raffinate I and raffinate II, MTBE, and MEK.
- The *Ethylene Annual Newsletter* covers various ethylene-related substances, including ethylene glycol, ethylene oxide, vinyl chloride monomer, HDPE, LDPE, and LLDPE.
- The *Light Hydrocarbon Feedstocks Newsletter* is primarily concerned with ethane, propane (including LPG), and the butanes, but also examines developments for the heavy oil feedstocks, natural gas, naphtha, gas oil, and crude.
- The *Methanol Annual Newsletter* concentrates upon methyl alcohol.
- The *Propylene Annual Newsletter* examines propylene oxide, propylene glycol, propylene alkylate, acrilonitrile, polypropylene, cumene and isopropyl alcohol, and propylene itself.
- The *Toluene/Xylenes Annual Newsletter* analyses trends relevant to the aromatics other than benzene, including toluene, mixed and solvent xylenes, orthoxylene, paraxylene, phthalic anhydride, DMT, TPA, and TDL.

Each of the newsletters has a corresponding price forecast, available separately, which projects current trends for the respective products up to two or more years in advance. At the time of writing, forecasts are extended through the fourth quarter of 1985. Every forecast appears with an explicit explanation of how DeWitt determines the anticipated course of prices.

In order to access the DeWitt newsletters, you must subscribe to those newsletters of interest to you. Applicable INFOMAGIC rates in North America for access to the DeWitt newsletters are $1.50 per minute at 300 baud and $3.00 per minute at 1200 baud. For subscription information or detailed answers to questions concerning DeWitt, please contact:

DeWitt & Company
3650 Dresser Tower
601 Jefferson Street
Houston, Texas 77002

Telephone: 713 652–0576
Telex: 762854 HOU

or the I. P. Sharp MAILBOX group *OILQ*. ■

Ken Bell, Houston

1982CFD Data Base Targets Markets for Financial Services

SRI International (formerly the Stanford Research Institute) is a non-profit research and consulting organization providing a variety of services to business and government. SRI has made available for the first time its 1982 Consumer Financial Decisions (CFD) data base through the I. P. Sharp timesharing service. This data base comprises the results of SRI's comprehensive survey of consumers' use of financial products: why they buy certain products, how they buy them, and how frequently they buy them. It is unique both in its size and in the scope of its coverage of demographics, financial holdings, and the use of and attitudes toward financial products, institutions, and methods of delivery.

The need for CFD clients to get rapid, meaningful answers to market research questions led SRI and I. P. Sharp to develop this proprietary data base on the I. P. Sharp timesharing service. With a local telephone call to the I. P. Sharp Network, users can generate survey crosstabulation reports in minutes using I. P. Sharp's software package XTABS.

The data base aids the user in obtaining immediate answers to specific problems and opportunities. For example:

- A credit union manager needs information on ATM use among credit union customers. By typing a single command into the terminal, he can pull the required information from the 1982CFD data base in a few minutes.
- An insurance company is considering offering financial planning as a separate service. Using 1982CFD, the market research department can provide immediate information on current use of advisory services and attitudes toward financial advice among insurance purchasers.
- A strategic planner for a bank needs to report on the potential market for a new product at a meeting later that day. Using 1982CFD, she can run tables that tell her what customer segments are high users of similar products, and whether these segments are likely to purchase the new product from the bank.
- A large financial firm is debating whether to purchase a discount brokerage firm. With 1982CFD data, it finds that interest in discount brokerage is low among its current customer base, but that current holdings and responses to questions about pricing indicate that interest could be developed among its customers.

Used as a decision-making resource by financial institutions such as banks, insurance companies, and brokerage firms, the 1982CFD data base provides subscribers with a major competitive advantage by giving them a better understanding of the consumer marketplace and how to profit in that environment. The data base comprises over 1900 variables for approximately 3500 households, weighted to represent the total U.S. population. Users can select from hundreds of demographic samples. There are over 28 million facts which can be segmented, along with comparisons and percentages. Because of its unique scope, the 1982CFD data base is the most extensive and authoritative source available for information on consumer needs for financial services.

Users can produce crosstabulations like the following example with only one or two commands. Each crosstabulation contains cell counts detailing information for each combination of the questionnaire responses and demographics. Within each cell are **frequencies** (the number of households in a certain demographic category who answered a certain way) and **percentages** (the percentage of the cell's frequency compared to the total demographic group and total households answering that same question in the same way).

To produce the crosstabulations, each client receives a user's guide, a list of demographic variables, and a copy of the survey questionnaire that has been coded for easy reference. The user needs only to designate the question and the desired demographics or market segments, and a complete crosstabulation is produced within seconds.

For additional information or a demonstration of the 1982CFD data base, please contact:

. . . 1982CFD Data Base

Catherine Sheldon or Wendy Abbott
Consumer Financial Decisions
SRI International
333 Ravenswood Avenue
Menlo Park, California 94025

Telephone: 415 859–3659 or 415 859–5709

If you have survey data that you are interested in interpreting and analysing, contact your I. P. Sharp representative for more information about XTABS. ■

Catherine Sheldon and Maria Martin, SRI International, Menlo Park, California

DATE: DEC 1, 1983 PAGE 1
REPORT 'CURRENT' SRI 1982 CFD DATA BASE

WHICH OF THE FOLLOWING HAVE YOU
DONE IN THE PAST TWO YEARS?

TABULATE 'P83Q8A' BY 'CFD11,CFD12,CFD13,CFD14,CFD17,CFD18'

	ROW TOTAL	TOP 3PCT OF INCOME	TOP 3PCT OF NET WORTH	INCOME 15K + BUSI-NESS OWNER	INCOME 15K + ELITE OCCU-PATION	YOUNG DUAL EAR-NERS	YOUNG AFFL UENT
WTD BASE	8873	271	268	599	459	755	720
TOTAL	7620	256	252	532	420	612	604
RESPONDING	85.9	94.2	94.2	88.8	91.5	81.0	84.0
	100.0	3.4	3.3	7.0	5.5	8.0	7.9
FUNDS FROM	1509	150	138	177	184	104	182
SAVINGS TO	17.0	55.4	51.5	29.6	40.0	13.8	25.4
MONEY MARKET	100.0	10.0	9.1	11.7	12.2	6.9	12.1
OPEN IRA	1097	123	84	145	147	70	98
	12.4	45.4	31.5	24.3	31.9	9.3	13.6
	100.0	11.2	7.7	13.3	13.4	6.4	8.9
REPLACE	297	19	18	43	21	51	76
WHOLE LIFE	3.3	6.8	6.6	7.1	4.5	6.8	10.5
W/ TERM INS.	100.0	6.3	6.0	14.4	7.0	17.2	25.5
TAKE OUT	627	65	60	91	71	63	53
LOAN AGAINST	7.1	24.1	22.3	15.2	15.5	8.4	7.3
INS. POLICY	100.0	10.4	9.5	14.5	11.4	10.1	8.4
CLOSE VISA/	1333	78	65	130	125	116	179
MASTERCARD	15.0	28.9	24.3	21.8	27.2	15.3	24.9
ACCT(S)	100.0	5.9	4.9	9.8	9.4	8.7	13.4
BUY ADDI-	921	54	37	110	77	123	117
TIONAL LIFE	10.4	20.0	13.8	18.3	16.8	16.3	16.2
INSURANCE	100.0	5.9	4.0	11.9	8.4	13.4	12.7
MOVE CHECKNG	1319	64	64	132	101	152	174
ACCT TO NEW	14.9	23.7	23.8	22.0	22.0	20.1	24.2
INSTITUTION	100.0	4.9	4.8	10.0	7.6	11.5	13.2
BEGIN USING	187	22	17	23	37	20	27
DISCT BROKER	2.1	8.2	6.5	3.9	8.1	2.7	3.8
	100.0	12.0	9.3	12.4	19.9	10.9	14.5
SWITCH AUTO	984	32	28	72	57	128	166
INS POLICIES	11.1	11.7	10.4	12.0	12.5	16.9	23.1
	100.0	3.2	2.8	7.3	5.8	13.0	16.9
GET 2ND	228	17	9	46	25	43	35
MORTGGE LOAN	2.6	6.4	3.2	7.6	5.3	5.7	4.9
	100.0	7.7	3.8	20.0	10.8	18.8	15.4
NONE OF	3261	28	45	101	68	225	133
THESE	36.7	10.1	17.0	16.8	14.9	29.8	18.4
	100.0	0.8	1.4	3.1	2.1	6.9	4.1

PERCENTAGES OF: WTD BASE, ROW TOTAL

I. P. Sharp's First Distributed Software Symposium

Seventy-five representatives from 34 SHARP APL sites met in Toronto, from 14 to 17 November, at the King Edward Hotel. Since there are SHARP APL systems installed in data centres in Europe, the United Kingdom, the United States, Canada, the Far East, Australia, and South Africa, this symposium provided the oportunity for representatives to share information and to discuss development plans.

During the first two days, sessions concentrated on the concerns of management from data processing departments, and application software and system software groups. I. P. Sharp personnel and site representatives covered a variety of topics, which included:

- the integration of the personal computer into the information centre
- interfaces to data base management systems
- APL as a tool for managing the data centre
- SHARP APL system developments, including SHARP APL running under MVS/XA, a new 3270 session manager, transaction billing, and national language alphabetics
- the I. P. Sharp Network for linking the I. P. Sharp timesharing service to distributed SHARP APL systems, and to non-SHARP APL systems
- recent and current developments to the X25 and SDLC protocols
- application software developments

One of the most valuable sessions of the symposium was the panel discussion on users' concerns. Here site reps exchanged ideas on a variety of topics such as implementation methods, special applications, and the use of technologies. The discussion gave users the opportunity to air their concerns to I. P. Sharp personnel in the areas they felt were most crucial for development efforts and to indicate application software products most in demand. As a result of this session, the SHARP APL Distribution and Development Groups are revising their priorities to meet users' concerns and are developing solutions for the problem areas.

On the final day of the symposium, the System Programmers' Workshop on System Tuning covered the problems encountered in tuning various system and environmental parameters to improve SHARP APL system performance and to relieve bottlenecks.

All the symposium attendees agreed that the past three days had been very worthwhile, and they look forward to future meetings. ■

1984 Users Meeting
15-17 October 1984
The Westin Hotel, Toronto

Rapid changes in computing technology and the widespread use of personal computers are presenting new challenges to the business environment and to everyone in our industry. At the 1984 Users Meeting you will hear how others are meeting these new advances in sessions focusing on personal computers, fourth generation languages, and information centres.

The program for the Users Meeting will be an integrated one, taking into account the new directions that I. P. Sharp as a company is headed, as well as the new directions of our users. We haven't finalized the program yet, but the 1984 Users Meeting promises to be a conference that you won't want to miss. Circle 15-17 October on your calendar now. ■

publications

Bookends

MABRA Users' Guide *(revised)*

MABRA is I. P. Sharp's online data management system. Companies use MABRA to collect and maintain their personnel data, detailed information on customers and distributors, and inventory data. Whatever data you have that is routine and very structured may be stored on a MABRA system.

This revised users' guide now includes all the changes to MABRA since the 1980 edition of MABRA. Among the changes are the report-writing features, such as the Generate Report command, the commands to specify preprinted forms or letterhead, and a more convenient way to validate records.
October 1983, 109 pp., $8.00

Federal Reserve Board Weekly Data Base Manual *(new)*

The Federal Reserve Board Weekly (FRBW) data base contains banking and monetary statistics released weekly by the Board of Governors of the U.S. Federal Reserve System. The data base includes 189 time series, most of which begin in 1982. Banks, financial institutions, and companies involved in the U.S. banking industry will find the data useful for economic studies, forecasting, business analysis, and market studies.
December 1983, 18 pp., $3.00

MICROCOMM Users' Manual *(revised)*

MICROCOMM is a microcomputer product which allows IBM Personal Computer users to communicate with the I. P. Sharp Network. Using MICROCOMM, you can quickly and conveniently sign on to the I. P. Sharp timesharing service to access information in the I. P. Sharp public data bases, and transfer data between the I. P. Sharp system and IBM PC files.

MICROCOMM Version 2.0 includes these new features: the ability to display the APL character font (by replacing the character ROM); 12 pages of video memory, with line-by-line or page-by-page scrolling; automatic handling of Hayes (or Hayes-compatible) Smartmodems; and the ability to view files on a diskette or on hard disk from within MICROCOMM.
November 1983, 32 pp. + diskette, $50.00

mailing request

- ☐ Please amend my mailing address as indicated.
- ☐ Please add the following name(s) to your Newsletter mailing list.
- ☐ Please send me a Publications Order Form.
- ☐ Please add my name to the Aviation Newsletter mailing list.
- ☐ Please add my name to the Energy Newsletter mailing list.
- ☐ Please add my name to the Financial and Economic Newsletter mailing list.
- ☐ Please add my name to the PROMIS Newsletter mailing list. PROMIS is a computer-aided manufacturing system for microelectronics.

Name: _____

Title: _____

Company: _____

Address: _____

_____

_____

☐ Please send me information about your courses in:

City: _____ Tel.: _____

The *I.P. Sharp Newsletter* is published by I.P. Sharp Associates, Suite 1900, 2 First Canadian Place, Toronto, Canada M5X 1E3.
Your comments and contributions are welcome.

Editor: Irene Shimoda
Design and Layout: Susan Storey, Carol Rutherford
Circulation: Mary Kopfensteiner

Printed in Canada
January 1984
ISSN 0226 854X

international offices

Aberdeen
I.P. Sharp Associates Limited
5 Bon Accord Crescent
Aberdeen AB1 2DH
Scotland
(0224) 25298

Amsterdam
InterSystems BV
Kabelweg 47
1014 BA Amsterdam
The Netherlands
(020) 86 80 11
Telex: 18795 ITS NL

Atlanta
I.P. Sharp Associates, Inc.
1210 S. Omni International
Atlanta, Georgia 30335
(404) 586-9600

Boston
I.P. Sharp Associates, Inc.
1 Liberty Square
Boston, Massachusetts 02109
(617) 542-2313

Brisbane
I.P. Sharp Associates Pty. Ltd.
6th Floor, Dalgety House
79 Eagle Street
Brisbane, Queensland 4000
Australia
(07) 229-8330

Brussels
I.P. Sharp Europe SA
Boulevard de la Cambre 36, bte 5
1050 Bruxelles
Belgium
(02) 649 99 77

Calgary
I.P. Sharp Associates Limited
Suite 550, Bow Valley Square 4
250-6th Avenue S.W.
Calgary, Alberta T2P 3H7
(403) 265-7730

Canberra
I.P. Sharp Associates Pty. Ltd.
16 National Circuit
Barton, A.C.T. 2600
Australia
(062) 73-3700

Chicago
I.P. Sharp Associates, Inc.
Suite 3860
55 West Monroe Street
Chicago, Illinois 60603
(312) 782-3177

Copenhagen
I.P. Sharp ApS
Østergade 24B
1100 Copenhagen K
Denmark
(01) 11 24 34

Coventry
I.P. Sharp Associates Limited
7th Floor B Block
Coventry Point, Market Way
Coventry CV1 1EA England
(0203) 56562

Dallas
I.P. Sharp Associates, Inc.
Suite 1148, Campbell Center
8350 North Central Expressway
Dallas, Texas 75206
(214) 369-1131

Denver
I.P. Sharp Associates, Inc.
Suite 416
5680 South Syracuse Circle
Englewood, Colorado 80111
(303) 741-4404

Dublin
I.P. Sharp Associates Limited
Segrave House
Earlsfort Terrace
Dublin 2, Ireland
(01) 763605

Düsseldorf
I.P. Sharp GmbH
Kaiserswertherstrasse 115
4000 Dusseldorf 30
West Germany
(0211) 45 20 52

Edmonton
I.P. Sharp Associates Limited
2358 Principal Plaza
10303 Jasper Avenue
Edmonton, Alberta T5J 3N6
(403) 428-6744

Frankfurt
I.P. Sharp GmbH
Rossertstrasse 15
am Palmengarten
6000 Frankfurt 1
West Germany
(0611) 72 47 73

Halifax
I.P. Sharp Associates Limited
Suite 706, Cogswell Tower
2000 Barrington Street
Halifax, Nova Scotia
B3J 3K1
(902) 423-6251

Helsinki
TMT-Team Oy (Agent)
Westendintie 1
SF-02160
Espoo 16, Finland
(0) 452 1155

Hong Kong
I.P. Sharp Associates (HK) Limited
Suite 606, Tower 1
Admiralty Centre, Hong Kong
5-294341

Houston
I.P. Sharp Associates, Inc.
Suite 375, One Corporate Square
2600 Southwest Freeway
Houston, Texas 77098
(713) 526-5275

London, Canada
I.P. Sharp Associates Limited
Suite 510, 220 Dundas Street
London, Ontario N6A 1H3
(519) 673-4060

London, England
(European Headquarters)
I.P. Sharp Associates Limited
132 Buckingham Palace Road
London SW1W 9SA, England
(01) 730-4567
Telex: 8954178 SHARP G

Los Angeles
I.P. Sharp Associates, Inc.
Suite 1230
1801 Century Park East
Los Angeles, Ca. 90067
(213) 277-3878

Madrid
I.P. Sharp Associates Limited
Serrano 23, Piso 8
Madrid 1, Spain
(91) 276 70 54

Melbourne
I.P. Sharp Associates Pty. Ltd.
520 Collins St., 7th Floor
Melbourne, Victoria
3000, Australia
(03) 614-1766

Mexico City
Teleinformatica de Mexico SA
(Agent) Mail to:
Arenal N 40, Chimalistac
Mexico 20 D.F., Mexico
(905) 550-8033

Miami
I.P. Sharp Associates, Inc.
Suite 240
15327 N.W. 60th Avenue
Miami Lakes, Florida 33014
(305) 556-0577

Milan
Informatical Society Italia Srl
(Agent)
Via Eustachi 11
20129 Milan, Italy
(02) 221 612

Montreal
I.P. Sharp Associates Limited
Suite 1610
555 Dorchester Boulevard W.
Montreal, Quebec H2Z 1B1
(514) 866-4981

New York
I.P. Sharp Associates, Inc.
Suite 210
230 Park Avenue
New York, N.Y. 10169
(212) 557-7900

Newport Beach
I.P. Sharp Associates, Inc.
Suite 1135
610 Newport Center Drive
Newport Beach, Ca. 92660
(714) 644-5112

Oslo
I.P. Sharp A/S
Postboks 486 Sentrum
Dronningens gate 34
Oslo 1, Norway
(02) 41 17 04

Ottawa
I.P. Sharp Associates Limited
Suite 600, 265 Carling Ave.
Ottawa, Ontario K1S 2E1
(613) 236-9942

Palo Alto
I.P. Sharp Associates, Inc.
Suite 201, 220 California Ave.
Palo Alto, Ca. 94306-1683
(415) 327-1700

Paris
I.P. Sharp SARL
9 Rue du Cirque
75008 Paris
France
(1) 225 98 20

Rochester
(United States Headquarters)
I.P. Sharp Associates, Inc.
1200 First Federal Plaza
Rochester, N.Y. 14614
(716) 546-7270

Rome
Informatical Society Italia Srl
(Agent)
Piazza Della Rotonda 2
00100 Rome, Italy
(06) 656-5925

San Francisco
I.P. Sharp Associates, Inc.
Suite C-415, 900 North Point St.
San Francisco, Ca. 94109
(415) 673-4930

Saskatoon
I.P. Sharp Associates Limited
Suite 303, Financial Bldg.
230-22nd St. E.
Saskatoon, Saskatchewan
S7K 0E9
(306) 664-4480

Seattle
I.P. Sharp Associates, Inc.
Suite 223, Executive Plaza East
12835 Bellevue Redmond Road
Bellevue, Washington 98005
(206) 453-1661

Seoul
Daewoo Corporation (Agent)
541, 5-Ga, Namdaemoon-Ro
Jung-Gu (CPO Box 2810)
8269 Seoul, Korea.
771-91/2
Telex: 23341-5/24295 DAEWOO K

Singapore (Far East H.Q.)
I.P. Sharp Associates (S) Pte.Ltd.
77 Robinson Road, No. 14-00
SIA Building
Singapore 0106
Republic of Singapore
2230211
Telex: 20597 IPSAS RS

Singapore
Singapore International
Software Services Pte. Ltd.
(Agent)
77 Robinson Road, No. 14-00
SIA Building
Singapore 0106
Republic of Singapore
2230211

Stockholm
I.P. Sharp AB
Kungsgatan 65
S111 22 Stockholm, Sweden
(08) 24 43 90

Sydney (Australian H.Q.)
I.P. Sharp Associates Pty. Ltd.
8th Floor, Carlton Centre
55 Elizabeth St.
Sydney, New South Wales 2000
Australia
(02) 232-6366

Tokyo
INTEC, Inc. (Agent)
Dai 23 Mori Bldg. 4th Floor
1-23-7 Toranomon
Minato-ku, Tokyo 105
Japan
(03) 502 8781
Telex: 2322008 INTECA J

Toronto
(International Headquarters)
I.P. Sharp Associates Limited
Suite 1900
2 First Canadian Place
Toronto, Ontario
M5X 1E3
(416) 364-5361
Telex: 0622259 I P SHARP TOR

Vancouver
I.P. Sharp Associates Limited
Suite 902, 700 West Pender St.
Vancouver, B.C. V6C 1G8
(604) 687-8991

Victoria
I.P. Sharp Associates Limited
Chancery Court
1218 Langley Street
Victoria, B.C. V8W 1W2
(604) 388-6365

Vienna
I.P. Sharp Ges.mbH
Renngasse 4
A-1010 Wien, Austria
(0222) 66 42 48

Warrington
I.P. Sharp Associates Limited
1-3 Dolmans Lane
Warrington, Cheshire
WA1 2ED England
(0925) 50413/4

Washington
I.P. Sharp Associates, Inc.
Suite 305, 2033 K Street N.W.
Washington, D.C. 20006
(202) 293-2915

Wayne
I.P. Sharp Associates, Inc.
Suite 303
155 Willowbrook Blvd.
Wayne, New Jersey 07470
(201) 785-8050

White Plains
I.P. Sharp Associates, Inc.
Suite 312 West
701 Westchester Avenue
White Plains, New York 10604
(914) 328-8520

Winnipeg
I.P. Sharp Associates Limited
Suite 208
213 Notre Dame Avenue
Winnipeg, Manitoba R3B 1N3
(204) 947-1241

Zurich
I.P. Sharp AG
Fortunagasse 15
CH-8001 Zurich, Switzerland
(01) 211 84 24

I. P. Sharp Network

APL OPERATOR VOICE 416 363-2051 COMMUNICATIONS 416 363-1832

Our private, packet-switched network connects with the Value Added Networks—Datapac, Datex-P, PSS, Telenet, Telepac, Transpac, and Tymnet—to provide access from the 46 countries listed below:

• Argentina • Australia • Austria • Bahamas • Bahrain • Barbados • Belgium • Bermuda • Brazil • Canada • Chile • Denmark • Dominican Republic • Finland • France • French Antilles • Germany • Great Britain • Hong Kong • Hungary • Ireland • Israel • Italy • Japan • Korea • Kuwait • Luxembourg • Mexico • The Netherlands • New Zealand • Norway • The Philippines • Portugal • Puerto Rico • Qatar • Republic of Singapore • Republic of South Africa • Saudi Arabia • Scotland • Spain • Sweden • Switzerland • Taiwan • United Arab Emirates • U.S.A. • Virgin Islands

SHARP APL is accessible from over 600 places via a local telephone call. Please ask your nearest I.P. Sharp office or representative for a complete list of access points and access procedures. Our private network also connects with the worldwide Telex network via the Amsterdam and Rochester, New York nodes.

JANUARY/FEBRUARY 1984
Volume 12 Number 1

I.P. Sharp newsletter

technical supplement 48

Making Things Fit, Part 3

How big should a file component be?

The question of how big a file component should be really addresses two issues:

1) What is an optimal size?
2) How can data be segmented to achieve this optimal size?

An important factor in answering this question is that the cost of reading a component grows more slowly than the size of the component. For example, the CPU time required to read 100 components containing ⍳5000 is only three times the time required to read the same number of components containing ⍳500, even though ten times as much data is being moved. Thus, we learn that using *large components* instead of small ones can reduce execution time.

The size of the active workspace is the real limiting factor on the size of file components. A good rule of thumb (and by no means infallible) is: your method of segmenting data should ensure that no component exceeds ten per cent of the size of a clear workspace. This means that in a 350K workspace, no component should ever exceed 35K bytes. This leaves a reasonable amount of room for other variables and functions in the workspace. While ten per cent may be the maximum, five per cent is a more reasonable average. This allows for some growth, which is important.

The two major ways to split data into file components correspond to the two types of data—uniform and non-uniform. With uniform data, every scalar in an array represents a fact. These facts are numbers, and they all have the same unit of measure. For example, a 12-element vector of monthly budget figures in dollars is uniform data.

With non-uniform data, a collection of scalars represents a fact. The scalars can be numbers or characters. The numbers may be measurable quantities or identifiers. The scalars may be grouped into units of meaning called **fields**, and an entire fact is called a **record**. This type of data is non-uniform because the data types of the scalars may differ, as well as the unit of measure (if any) of the numbers. For example, a vector consisting of a department number, account number, dollar amount, and date is a record of non-uniform data.

How do you segment uniform data? Consider a typical APL application. A company wants to monitor the performance of its sales organization. It has monthly statistics by sales unit by product for several years, containing both the sales plan and the actual figures. This can be treated as a five-dimensional array: Year by Type (plan, actual) by Unit by Product by Month. If there are a lot of units or products, the data won't fit in the workspace.

To make the data fit, you must split off some of the dimensions. If there are 50 units and 200 products, you might make each file component a matrix in which the rows represent products, and the columns represent months. Each matrix will be 200 by 12. If the data is stored in four-byte integers, the component will take approximately 9600 bytes, which is a reasonable size. There will be a matrix for each Year/Type/Unit combination. For five years of data, that will mean 5×2×50 = 500 components.

To locate the component associated with a

given combination, you can compute the following:

```
1+(0,NUMBER∆OF∆TYPES,NUMBER∆OF∆UNITS)⊥
  ¯1+YEAR∆INDEX,TYPE∆INDEX,UNIT∆INDEX
```

You may want to provide for one or more of the dimensions becoming longer (a very likely possibility). In this case, use the number computed above to index into a vector that indicates which component relates to that combination. When a dimension becomes longer, you will have to insert elements in the vector to represent the new data, but the components on file can remain where they are.

There are no hard-and-fast rules for splitting dimensions to obtain an optimal file design. Listed below are some helpful guidelines:

1) Keep data that will be used together most frequently in the same file component to minimize file operations
2) Split as few dimensions as possible to minimize looping
3) Provide an orderly means for growth to minimize maintenance headaches

How can non-uniform data be segmented? In *Technical Supplements 40* and *41*, six methods of searching files containing such data are described:

1) Sequential search of a random file
2) Sequential search of an inverted file
3) Binary search of a sorted file
4) Directory search of a sorted file
5) Sequential search of a buffered (random or sorted) file
6) Algorithmic search of a hashed file

These six methods reduce to four file organizations: random, sorted, inverted, and hashed. The following discussion explains how to group records under each scheme in order to prevent *WS FULL* and maximize performance.

Managing a file in which the records are not stored in any special order is not very hard. Records are blocked together in components. The component size should be between five and ten per cent of □*WA* in a clear workspace. New records are appended to the bottom of the last component. When it is filled, a new component is appended to the end of the file. The following function shows how this works.

```
    ∇ RECORD APPEND∆RANDOM TIE
      ;BLOCK;PTR
[1]   PTR←¯1+1↑1↓□SIZE TIE
[2]   BLOCK←□READ TIE,PTR
[3]   →((1↑ρBLOCK)<MAXBLOCKSIZE)ρAPPEND
[4]   BLOCK←(0,¯1↑ρRECORD)ρ0
[5]   PTR←PTR+1
[6]   BLOCK □APPEND TIE
[7] APPEND:BLOCK←BLOCK,[1] RECORD
[8]   BLOCK □REPLACE TIE,PTR ∇
```

Managing a file of sorted records takes a bit more effort. If the file is permanently sorted, a directory can be created to find records more easily. Records are blocked together in components with the other records that have the same identifying field (sort key). But since the distribution of records over the different keys may not be even (and probably isn't), a means of segmenting the data further must be used.

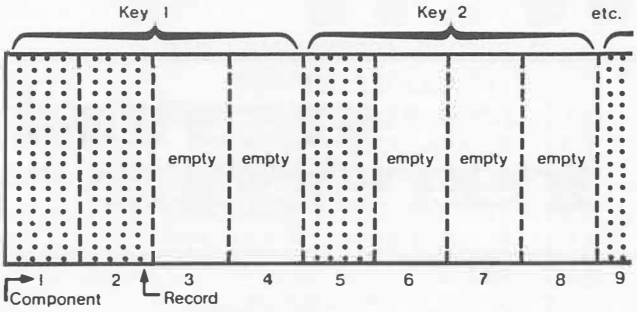
The solution is to allow for multiple components for each key. The directory must contain three columns:

1) the identifying field
2) the first component containing records that are identified
3) the last component containing records for that identifier.

Thus records are first split by key, and then again into blocks. The maximum block size should be five to ten per cent of □*WA* in a clear workspace, divided by the number of bytes each record needs.

A group of components is reserved for each key. The number of components reserved should be several times greater than the number of components you expect a single key to require. For example, if the average key will require one block of records, allocate five or more components per key. The extra components are reserved by appending empty arrays. The empty components aren't read unless they are needed, and so cost nothing. Empty components reserved for future use are a cheap way to allow for growth. The following diagram shows the layout of such a file structure.

. . . Making Things Fit



SORTED RECORDS BLOCKED BY RECORD

The following program shows how a new record is appended to such a file structure. Note that it attempts to place the record in the first block which isn't full, since deletions may have eliminated records. It also contains a paranoia check, just in case all your extra empty components have been used up.

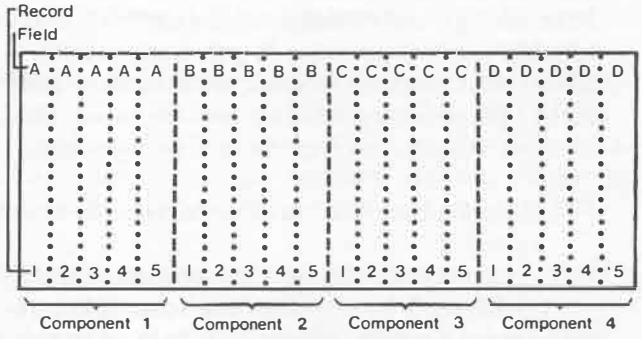
```
    ∇ RECORD APPENDΔDIRECT TIE
      ;BLOCK;ROW;FIRST;LAST
[1]   ROW←INDEX[;1]⍳RECORD[KEYFIELD]
[2]   CTR←INDEX[ROW;2]-1
[3]   LMT←INDEX[ROW;3]
[4]  LOOP:→(LMT<CTR←CTR+1)⍴END
[5]   BLOCK←⎕READ TIE,CTR
[6]   →((1↑⍴BLOCK)≥MAXBLOCKSIZE)⍴LOOP
[7]   BLOCK←BLOCK,[1] RECORD
[8]   BLOCK ⎕REPLACE TIE,CTR
[9]   →EXIT
[10]END:→(CTR>BLOCKSPERKEY)⍴FULL
[11]  BLOCK←(1⌈¯2↑⍴RECORD)⍴RECORD
[12]  BLOCK ⎕REPLACE TIE,CTR
[13]  INDEX[ROW;3]←CTR
[14]  →EXIT
[15]FULL:'NO SPACE FOR KEY ',⍕INDEX[ROW;1]
[16]EXIT: ∇
```

An inverted file structure segments its data by field, rather than by record. The simplest inverted structure puts all of the values of one field from all records in a separate component. If there are several thousand records in the file, the components will get too big to fit in the workspace. So, the data is first split by field, and then each field is split into blocks. The block size should be five to ten per cent of ⎕WA in a clear workspace, divided by the number of bytes one field entry takes.

There are two ways to lay out an inverted file:

1) All components which are part of a single field can be stored contiguously. If this structure is chosen, room for growth can be provided by putting empty components at the end of every group of components which comprise a single field. Such an organization must be reorganized when all of the empty components are filled.
2) All components which contain a given set of records can be stored contiguously. If this structure is chosen, no special provision for adding records need be made. When a block is filled up, a new one is started. However, such an organization must be reorganized when a field is added or deleted.

The following diagram shows the layout of an inverted file.



INVERTED RECORDS BLOCKED BY FIELD

The next program shows how a new record is appended to an inverted file. Note that it has to do a ⎕READ and ⎕REPLACE for each field in the record. While inverted files are fast to search, they are slow to update.

```
    ∇ RECORD APPENDΔINVERT TIE
      ;FLDCTR;FLDLMT;BLKCTR;BLKLMT;BLOCK;BASE
[1]   FLDCTR←0
[2]   FLDLMT←⍴RECORD
[3] FLDLOOP:→(FLDLMT<FLDCTR←FLDCTR+1)⍴FLDLOOPEND
[4]   BASE←BLOCKSPERFIELD×¯1+FLDCTR
[5]   BLKCTR←0
[6]   BLKLMT←BLOCKSPERFIELD
[7] BLKLOOP:→(BLKLMT<BLKCTR←BLKCTR+1)⍴BLKLOOPEND
[8]   BLOCK←⎕READ TIE,BASE+BLKCTR
[9]   BLOCK←BLOCK,RECORD[FLDCTR]
[10] BLOCK ⎕REPLACE TIE,BASE+BLKCTR
[11] →BLKLOOP
[12]BLKLOOPEND:→FLDLOOP
[13]FLDLOOPEND: ∇
```

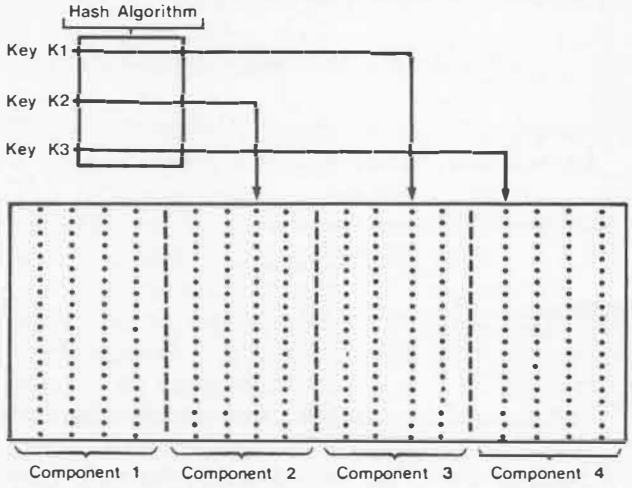
A hashed file contains non-uniform data stored by record. The records are neither in random order, nor sorted order. Instead, they

are scattered among a limited set of components based on an arithmetical calculation done on the identifying field. The calculation should distribute the records over the available components relatively evenly. A maximum block size should be established as with the sorted file.

What happens when too many records map into the same component? There is a fixed number of components used for the hashed records, and any beyond them are used as an overflow area. Records which don't fit in the component they are mapped to are stored in an overflow area, in sequential order. Obviously, if a significant portion of the records go to the overflow area, this method will be even slower than the sequential method to update and to search. The key to this method is to have enough components available, with a reasonable number of records per component. Since most hashing is done with modulo arithmetic, properly segmenting records using this method usually involves the following steps:

1) Approximate the number of records to be stored
2) Approximate the number of components to be used by dividing the total space required to store all records by a reasonable component size
3) Find a hash number (usually prime) which distributes the anticipated records relatively evenly over the components

The next diagram shows how records are mapped into components under this scheme.



HASHED RECORDS

The following program shows how a record is appended to a hashed file.

```
    ∇ RECORD APPENDΔHASH TIE
      ;BLOCK;CTR;LMT;PTR
[1]   PTR←1+HASHNUM|RECORD[KEYFIELD]
[2]   BLOCK←⎕READ TIE,PTR
[3]   →(MAXBLOCKSIZE≤1↑ρBLOCK)ρOVERFLOW
[4]   BLOCK←BLOCK,[1] RECORD
[5]   BLOCK ⎕REPLACE TIE,PTR
[6]   →EXIT
[7] OVERFLOW:CTR←HASHNUM
[8]   LMT←¯1+1↑1↓⎕SIZE TIE
[9] LOOP:→(LMT<CTR←CTR+1)ρMORE
[10]  BLOCK←⎕READ TIE,CTR
[11]  →(MAXBLOCKSIZE≤1↑ρBLOCK)ρLOOP
[12]  BLOCK←BLOCK,[1] RECORD
[13]  BLOCK ⎕REPLACE TIE,CTR
[14]  →EXIT
[15]MORE:BLOCK←(1⌈¯2↑ρRECORD)ρRECORD
[16]  BLOCK ⎕APPEND TIE
[17]EXIT: ∇
```

To decide which of these four methods to use in segmenting non-uniform data depends on the way the data will be used. The last three can all run quite efficiently under the proper circumstances. The suggestions in this section will make it possible for you to segment the data so that it fits in the active workspace.

Conclusion

The chief way to make data fit in a file is to design a flexible file structure. Such a flexible structure provides room for growth in the volume of data stored, as well as for variations in the distribution of the data over the identifiers. If you design files with these goals in mind, you will provide applications which serve their users well.

Acknowledgements

My thanks to J. Henri Schueler, Peter Wooster, and Gordon Ross for their help with this article. ■

Robert Metzger, Rochester